#21 Draft 4-25-03

Please type a plus sign (+) inside this box → [+]

PTO/SB/21 (08-00)
Approved for use through 10/31/2002. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	09/853,753
Filing Date	May 14, 2001
First Named Inventor	Bech-Hansen
Group Art Unit	1646
Examiner Name	Olga Chernyshev
Attorney Docket Number	45499.2

RECEIVED 03 APR 23 PM 2:59 USPTO/290

ENCLOSURES *(check all that apply)*

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment / Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [x] Certified Copy of Priority Document(s)
- [ ] Response to Missing Parts/ Incomplete Application
  - [ ] Response to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Assignment Papers *(for an Application)*
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
- [ ] After Allowance Communication to Group
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to Group *(Appeal Notice, Brief, Reply Brief)*
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) *(please identify below)*:

RETURN POSTCARD

CA 2,306,241

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm or Individual name	EDWARD TOO 41,435
Signature	
Date	22 April 2003

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, Washington, DC 20231 on this date: [ ]

Typed or printed name			
Signature		Date	

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

Offic de la propri intelle tuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Offi e
An Agency of Industry Canada

Bureau canadien des brevets
Certification

Canadian Patent Office
Certification

La présente atteste que les documents ci-joints, dont la liste figure ci-dessous, sont des copies authentiques des documents déposés au Bureau des brevets.

This is to certify that the documents attached hereto and identified below are true copies of the documents on file in the Patent Office.

Specification and Drawings, as originally filed, with Application for Patent Serial No: **2,306,241**, on May 12, 2000, by **N. TORBEN BECH-HANSEN**, for "Retinal GPI-Anchored small Leucine-Rich Proteoglycan Gene".

Agent certificateur/Certifying Officer

January 16, 2003
Date

Canada

(CIPO 68)
04-09-02

OPIC CIPO

RETINAL GPI-ANCHORED SMALL LEUCINE-RICH PROTEOGLYCAN GENE

FIELD OF THE INVENTION

This present invention is related to a gene encoding a novel small leucine-rich proteoglycan gene. In particular, this invention relates to a mammalian gene encoding the gene herein referred to as *NALPN*, wherein mutations of *NALPN* may cause complete X-linked congenital stationary night blindness.

REFERENCES

The following references are cited in the application as numbers in brackets ([ ]) at the relevant portion of the application. Each of these references is incorporated herein by reference.

1. Rodieck, R.W. *The First Steps in Seeing* (Sinauer, Sunderland, Massachusetts, 1998).

2. Rattner, *et al.* "Molecular genetics of human retinal disease". *Annu. Rev.Genet.* **33**, 85-131 (1999).

3. Carr "Congenital stationary night blindness." *Trans. Am. Ophthalmol. Soc.* **72**, 448-487 (1974).

4. Miyake *et al.* "Congenital stationary night blindness with negative eletroretinogram."

Arch. Ophthalmol. **104**: 1013-1020 (1986).

5. Sharp, *et al.* "Mechanisms and sites of loss of scotopic sensitivity: a clinical analysis of congenital stationary night blindness." *Clin. Vision Sci.* **5**: 217-230 (1990).

6. Boycott, *et al.* "Evidence for genetic heterogeneity in X-linked congenital stationary night blindness." *Am. J. Hum. Genet.* **62**:865-87 (1998).

7. Bech-Hansen, *et al.* "Localization of a gene for incomplete X-linked congenital stationary night blindness in Xp11.23 to the interval between DXS6849 and DXS8023." *Human Genetics* **103**:124-130 (1998).

8. Bech-Hansen., *et al.* "Loss-of-function mutations in a calcium-channel α_1-subunit in Xp11.23 cause incomplete X-linked congenital stationary night blindness. *Nature Genet.* **19**:264-267 (1998).

9. Strom, *et al.* (1998) "An L-type calcium-channel gene mutated in incomplete X-linked congenital stationary night blindness." *Nature Genet.* **19**:260-263 (1998).

10. Hocking, *et al.* "Leucine-rich repeat glycoproteins of the extracellular matrix.." *Matrix Biol.* **17**:1-19 (1998).

11. Medof, *et al.* "Cell-surface engineering with GPI-anchored proteins." *FASEB J.* **10**: 574-586 (1996).

12. Lewin, "*Genes*" (New York, Oxford University Press; 1997)

13. Kobe and Deisenhofer "The leucine-rich repeat: a versatile binding motif." *TIBS* **19**:415-421 (1994).

14. Plaas, *et al.* "Identification of the keratan sulfate attachment sites on bovine fibromodulin." *J. Biol. Chem.* **265**: 20634-2060 (1990).

15. Nielsen, *et al.*, "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites." *Protein Engineering* **10**: 1-6 (1997).

16. Englund "The structure and biosynthesis of glycosyl phosphatidylinositol protein anchors." *Annu. Rev. Biochem* **62**: 121-138 (1993).

17. Boycott, *et al.* "Integration of 101 DNA markers across human Xp11 using a panel of somatic cell hybrids." *Cell Cytogenet. Genet.* **76**: 223-228 (1997).

18. Fishman and Sokol "*Electrophysiologic Testing in Disorders of the Retina, Optic Nerve, and Visual Pathway.*" (Am. Acad. Ophthal., San Francisco, 1990)

19. Bergen, *et al.* "Localization of a novel X-linked congenital stationary night blindness locus: close linkage to the RP3 type retinitis pigmentosa gene region." *Human Molecular Genetics* 4:931-935 (1995).

20. Rowen and Koop *Automated DNA sequencing and analysis* (eds. Adams, M.D., Fields, C. & Venter, J.C.) 167-174 (Academic Press, London and San Diego, 1994).

21. Leimeister, *et al.*, "Developmental expression patterns of mouse sRFP genes encoding members of the secreted frizzled related protein family." *Mech. Dev.* **75**: 29-42 (1998).

22. Schaeren-Wiemers and Gerfin-Moser "A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells." *Histochemistry* **100**: 431-440 (1993).

23. Breitschopf, H. & Suchanek, G. *Nonradioactive in situ hybridization application manual.* pp. 136-140 (Boehringer Mannheim, Mannheim, Germany, 1996).

BACKGROUND

During mammalian retinal development a complex sequence of molecular events leads to the precise laminations and interconnections of the mature retina. In normal mature human retinas, rod and cone photoreceptors start the processing of vision, which proceeds through bipolar and ganglion cell retinal pathways to the brain [1]. Hereditary disease can perturb these

retinal pathways and cause either progressive degeneration or more stationary visual deficits [2]. Congenital stationary night blindness (CSNB) is a group of retinopathies that fall into the latter category of a selective retinal pathway disturbance that manifest at birth. CSNB has been recognized clinically for more than 100 years; genetic subtypes have been defined; and different sites of disease action have been postulated [3-5].

Patients with X-linked CSNB phenotypically exhibit normal fundi, but generally have reduced visual acuity, impaired night vision and, in addition, may exhibit myopia (or occasionally hyperopia), and nystagmus. Based on electroretinographic findings, patients with X-linked CSNB can have one of two forms of X-linked CSNB - complete or incomplete [4,6]. This clinical heterogeneity correlates with underlying genetic heterogeneity in which complete X-linked CSNB segregates with the CSNB1 locus in Xp11.4, and incomplete X-linked CSNB segregates with the CSNB2 locus in Xp11.23 [6,7]. Patients with incomplete X-linked CSNB who show both impaired rod and cone function were recently shown to have mutations in a voltage-gated L-type calcium channel α-$_{1F}$-subunit gene, *CACNA1* [8,9].

The biochemical defects underlying complete X-linked CSNB is unknown but may be revealed by identifying the gene involved in this disorder. The CSNB1 locus was previously reported to be on the proximal portion of the human X chromosome, between DXS556 and DXS8083, as described previously [6].

The identification of the gene which is causative of complete X-linked CSNB may allow

for development of diagnostic tests for this disorder and risk assessment in members of affected families. As well, identification of the gene that is causative of complete X-linked CSNB will provide information as to the basic defect in this retinal condition, which could lead to effective methods for treatment or cure of the disorder.

Leucine-rich repeat glycoproteins form part of the extracellular matrix (ECM) of mammalian cells [10]. The major components of the ECM are collagens, proteoglycans, glycosaminoglycans, fibronectin and, to a lesser extent, glycoproteins. These components are organized into a fibrillar meshwork, to provide mechanical strength and elasticity, and to create a structural framework that provides a substratum for cell adhesion and migration. The ECM plays an integral role in the pivotal processes of development, tissue repair, and metastasis. Within the ECM, the leucine-rich repeat glycoproteins are likely to perform more than a structural role, and also likely to be involved in regulating cell growth, adhesion and migration.

Many cell-surface proteins are anchored to the external surface of the plasma membrane by covalently attached glycoinositol phospholipids ("GPIs"). These anchors use a common structure as a general mechanism for membrane attachment, irrespective of protein function, and are added post-translationally at the time of the translocation of the protein across the endoplasmic reticulum [11]. Why certain proteins have GPI anchors is unknown.

The N-terminus of a secreted protein usually consists of a cleavable leader of 15-30 amino acids, which is called a signal sequence. The signal sequence is both necessary and sufficient for transfer of any attached polypeptide to the target membrane and is responsible for

directing ribosomes to attach to the endoplasmic reticulum as soon as the first few N-terminal amino acids are synthesized [12].

The identification of the gene, mutations of which cause incomplete X-linked CSNB1, will aid in the elucidation of the role of the protein in retinal function. Knowledge of the structure of this gene, from both naturally occurring mutations and engineered variants of the protein, will lead to studies of the structure-function relationships of the protein in the cellular environment and its role in the disease process. Further, the identification of the gene will provide a tool for the diagnosis of complete X-linked CSNB in individuals suspected of having this disorder.

SUMMARY OF THE INVENTION

The complete sequence of a gene on the short arm of the X-chromosome, herein referred to as *NALPN,* has been elucidated. *NALPN* has homology to members of the small leucine-rich proteoglycan family of genes, which is a subfamily of the leucine-rich repeat superfamily of proteins, and to GPI-anchored proteins. Mutational analysis of *NALPN* in 17 families with complete X-linked CSNB has identified 11 different mutations that are predicted to cause stop, deletion, insertion or missense mutations in the protein product of *NALPN,* herein referred to as Nyctalopin. Together, these findings establish that mutations in *NALPN* cause complete X-linked CSNB.

The present invention provides a mammalian nucleotide sequence encoding a novel small

leucine-rich proteoglycan expressed in the retina and also in the kidney. Thus, in one aspect, this invention is an isolated DNA molecule comprising a sequence of nucleotides that encodes a mammalian GPI-anchored small leucine-rich proteoglycan, including a human retinal and kidney GPI-anchored small leucine-rich proteoglycan.

In one embodiment, the invention comprises a DNA molecule that encodes a human retinal and kidney GPI-anchored small leucine-rich proteoglycan, and has a sequence of nucleotides selected from a group consisting of:

(a) the sequence set forth in Figure 7;

(b) a sequence of nucleotides that encodes the sequence of amino acids set forth in Figure 7.

In another embodiment, the invention comprises a DNA molecule that encodes a murine GPI-anchored small leucine-rich proteoglycan, which is the orthologue of the human retinal and kidney human small leucine-rich proteoglycan.

In another aspect, this invention comprises a substantially pure mammalian GPI-anchored small leucine-rich proteoglycan, including a human GPI-anchored small leucine-rich proteoglycan, represented by the sequence of amino acids set forth in Figure 7.

In another embodiment, the invention comprises a protein molecule that encodes a murine GPI-anchored small leucine-rich proteoglycan, which is the homologue of the human GPI-anchored human small leucine-rich proteoglycan.

In another aspect, this invention comprises an isolated RNA sequence that encodes a mammalian GPI-anchored small leucine-rich proteoglycan, including a human or murine GPI-anchored small leucine-rich proteoglycan. This invention also comprises an antisense RNA molecule having a sequence that is complementary to the mRNA encoding a mammalian GPI-anchored small leucine-rich proteoglycan, including a human or murine GPI-anchored small leucine-rich proteoglycan.

In another aspect, this invention comprises an expression vector, preferably a mammalian expression vector, comprising the nucleotide sequence of a mammalian GPI-anchored small leucine-rich proteoglycan, including a human and murine GPI-anchored small leucine-rich proteoglycan.

In another aspect, the invention comprises a method of diagnosing complete X-linked CSNB, which method includes screening for alterations in the sequence of nucleotides disclosed herein.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Diagnostic features of complete X-linked CSNB.

Figure 2: Minimal region for the CSNB1 locus. The chromosomal region from seven males that were analysed is indicated by the seven vertical lines. The thin black vertical line on the right

indicates the minimal genetic region established for the CSNB1 gene.

Figure 3: (a) Physical map of the CSNB1 minimal region indicating the location of overlapping BACs and PACs (short lines) and the chromosomal position of several genes in this region, including *NALPN*. The lower horizontal line demonstrates the genomic organization of *NALPN*, showing that it is comprised of three exons, with a translation start site in the second exon, a stop codon in the third exon and a polyadenylation sequence in the 3' untranslated region. (b) The amino acid sequence of Nyctalopin. The protein has 11 leucine-rich repeat motifs with a 24 amino acid consensus for small leucine-rich proteoglycans with cysteine clusters flanking the repeat core of the protein. The conserved amino acids are shown in bold.

Figure 4: cDNA expression profile of *NALPN* in various human tissues. (a) Upper panel shows the 755 bp fragment of the *NALPN* mRNA in retina and kidney tissue samples. Lower panel shows the 281 bp fragment of EST JRL4A1 which serves as a positive control. (b) *In situ* hybridization in human retinal sections showing the expression of *NALPN* in the outer- and inner-nuclear layers (ONL, INL), and in the ganglion cell layer (GCL).

Figure 5: Mutation analysis of the *NALPN* gene in families with complete X-linked CSNB. (a) Identification of a nonsense mutation in exon 3 (at nucleotide 1049) in patient 610-1 caused by a G to A transition, which changes a Trp to a stop codon. Segregation analysis of this mutation was performed by restriction endonuclease digestion. Affected individuals show the loss of a *Fok*I restriction site, female carriers have fragments indicating both the presence and absence of this restriction site, and unaffected males show only the *Fok*I site.

(b) Identification of a 24-bp deletion observed in six different families. This mutation results in the loss of eight amino acids beginning at codon 29.

(c) Segregation analysis of the 24-bp deletion demonstrated in panel (B). Genomic DNA was amplified by PCR and the products were subjected to agarose gel electrophoresis. The presence of the smaller PCR fragment representing the deletion was observed in affected males and carrier females.

(d) Identification of a missense mutation in patient P520-IV-27 caused by a T to A transversion at nucleotide 638, which changes a Leu codon to a Gln codon.

(e) Identification of an insertion mutation of 21 nt between nucleotides 444 and 445 in patient 650-1, which results in the addition of seven amino acids to the protein.

Figure 6: Summary of the 11 mutations of *NALPN* detected in 17 families with complete X-linked CSNB. Mutation refers to the position of the nucleotide changes, including insertions, deletions and changes. 'Codon change' shows the codons which have been changed by the mutations in *NALPN*.

Figure 7: Nucleotide sequence of human *NALPN* (SEQ ID NO: 1) with the amino acid sequence of Nyctalopin in single letter code underneath (SEQ ID No: 2).

DETAILED DESCRIPTION OF THE INVENTION

The details of the preferred embodiments of the present invention are set forth in the accompanying drawings and description below. Based on the details of the invention described

herein, numerous additional innovations and changes will become obvious to one skilled in the art.

A. Definitions

Unless otherwise indicated, all terms used herein have the same meaning as is commonly understood by one skilled in the art of the present invention. Practitioners are particularly directed to *Current Protocols in Molecular Biology* (Ausubel) or Maniatis *et al.*, Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory (1990), for terms of the art.

As used herein the following terms have the following meanings:

"**carrier**" refers to a female who is heterozygous for a single recessive gene and does not have the phenotype associated with complete X-linked CSNB.

"**expression vector**" refers to a recombinant DNA construct that comprises, among other elements, a DNA sequence of which expression is desired. An "expression vector" is used to introduce heterologous DNA into cells for expression of the heterologous DNA, as either an episomal element, or after incorporation into the cellular genome. An "expression vector" will contain all of the elements necessary for transcription of the DNA sequence functionally linked to the DNA sequence, including but not limited to a transcription initiation element, a transcription termination element, and elements that modulate expression of the DNA sequence, such as promoters or enhancers. These elements may be native to the DNA sequence of which expression is desired. An expression vector may contain elements that will regulate translation,

if translation of the resultant RNA transcript into a protein product is desired.

"**heterologous**" refers to DNA or RNA that does not occur, in nature, as part of the genome in which it is present, which is found in a location or locations in the genome that differ from that in which it occurs in nature, or which is present in the genome as a result of human manipulation of the genome. It is DNA or RNA that is not endogenous to the cell in which it is found, or that is endogenous to the cell but which has been manipulated *in vitro*, and has been artificially introduced into the cell. Heterologous DNA or RNA need not be incorporated into the host cell genome, but may be maintained episomally.

"**high stringency**" or "**conditions of high stringency**" means washing at low salt concentration, less than about 0.2 and preferably about 0.1 SSPE, and at high temperature, more than about 60°C and preferably about 65 °C. It will be understood that an equivalent stringency may be achieved by using alternative buffers, salts and temperatures.

"**Nyctalopin**" refers to the protein encoded by *NALPN,* and includes variants thereof which occur in nature or which can be generated experimentally. Such variants include proteins that have conservative amino acid substitutions, or those which have alterations in amino acid sequence that affect protein function.

"***NALPN***" unless indicated otherwise, "*NALPN*" refers to the gene that encodes Nyctalopin, and includes untranslated regions and regulatory and promoter sequences which influence the

expression of the gene and the translation of the transcript. In the appropriate context, *NALPN* may refer to the cDNA product of this gene.

"**precursor**" refers to a protein with the amino acid sequence corresponding to the sequence of the full length mRNA which, upon translation, results in a protein which may be further processed to form the mature Nyctalopin.

B. Mapping the Location of the Gene for Complete X-linked CSNB

The CSNB1 locus was previously reported to be on the proximal portion of the human X chromosome, between DXS556 and DXS8083 [6]. By analyzing additional families with complete X-linked CSNB, and by using new polymorphic markers developed on the basis of dinucleotide repeats within this minimal region, the minimal region can be further refined. Analysis of selected recombinant X chromosomes in the set of families with complete X-linked CSNB moves the distal boundary of the CSNB1 minimal region from DXS556 to the interval between 200L4CA1 and DXS8012 (Fig. 2 and Fig. 3(a)). From further analysis, the proximal crossover previously observed in patient V:1 in family P23 [19] limits the proximal boundary to between DXS1207 and DXS228 (Fig. 2 and Fig. 3(a)). Therefore, the minimal region for the CSNB1 locus is limited to the interval between 200L4CA1 and DXS228 in Xp11.4. The methods used for localizing genes on human chromosomes using these and other techniques are known to those skilled in the art.

C. Identification of a Candidate Gene for CSNB1

To position genetic markers accurately across the CSNB1 minimal region and identify

candidate genes for the CSNB1 locus, a robust physical map of the CSNB1 minimal region in Xp11.4 is developed. A subset of BAC and PAC clones from the minimal tiling path of the estimated 1.2 Mb CSNB1 minimal region is sequenced to between 1-2.7-fold redundancy. A sub-library is constructed [20] for each of the BAC clones and random clones from each sub-library are sequenced with the aid of ABI 373 or 377 sequencing machines and fluorescently labeled primers (ABI, Amersham). DNAStar software is used for gel trace analysis and contig assembly as well as DNA and protein alignments. DNA and protein sequences are then examined against available public databases using the various Blast programs available through the network server at the National Center for Biotechnology Information. A novel open reading frame is present in this region, which we have designated to be *NALPN*.

Candidates for the CSNB1 gene are expected to be expressed in the retina and located in the CSNB1 minimal region. Expression of *NALPN* is assessed by PCR amplification of a QUICK-Screen™ Human cDNA Library Panel, (Clontech) using primers which span exons 1 to 3. PCR products are electrophoresed through a 1% agarose gel and visualized by ethidium bromide staining. The 755-bp PCR product is detected in the retinal and kidney cDNA library (Figure 3). RNA *in situ* hybridization is performed as described [21-23], using a 668-bp antisense probe (from nt1557-2224 of the cDNA). These studies reveal that retinal expression of *NALPN* occurs predominantly in the inner- and outer-nuclear layers of the human retina but is also seen in some cells of the ganglion cell layer (Figure 4).

The extended *NALPN* cDNA sequence is established by sequencing of PCR and RACE products using first-strand cDNA from total human retinal RNA as the template for PCR.

Touchdown PCR using the Failsafe™ PCR Premix Selection Kit is carried out according to the manufacturer's protocol (Epicentre Technologies). RACE is carried out using Human Retina Marathon-ready cDNA (Clontech). A secondary amplification using a nested *NALPN* specific primer is performed. RACE and PCR products are gel purified using the Concert Gel Extraction kit (Life Technologies) and sequenced using the ThermoSequenase ^{33}P-radiolabelled terminator cycle sequencing kit (Amersham Life Science). To establish the genomic organization of the *NALPN* gene, the full-length cDNA sequence of *NALPN* is compared to the genomic sequence derived from our analysis of BAC clone 378P5 and that produced by the Sanger Centre for BAC clone 169I5.

The cDNA sequence of *NALPN* consists of a 1443 bp open reading frame that codes for a protein of 481 amino acids (Figure 7). The 5'-untranslated region is 98 bp long, and the translation initiation site lies in exon 2. A polyadenylation site lies 688 nt downstream from the stop codon of the open reading frame, and the 3'-untranslated region is at least 753 bp long. On the chromosome, *NALPN* is organized in three exons, spanning 28 kb of genomic sequence. The 3'-end lies 26 kb distal to the proximal end of PAC clone 169I5 (Figure 3). The open reading frame of *NALPN* is contained in exons 2 and 3. Intron 2 spans 25.5 kb and encompasses the marker DXS8012.

D. Characteristics of Nyctalopin

NALPN encodes a 481 amino acid protein, herein called Nyctalopin, which has sequence similarity with members of the superfamily of proteins containing tandem arrays of the leucine-rich repeat (LRR) motif [10,13]. Such proteins are known to function in protein-protein

interactions, especially in matrix assembly, and therefore Nyctalopin may possibly be mediating specific neural connections between cells in the retina. Moreover, the presence of the 24 amino acid consensus: x-x-I/V/L-x-x-x-x-F/P/L-x-x-L/P-x-x-L-x-x-L/I-x-L-x-x-N-x-I/L (where I,V,L,F,P and N are single letter amino acid codes and "x" represents any amino acid) in the core protein with cysteine clusters flanking the LRR domain (see Figure 3B), qualifies Nyctalopin as a new member of the subfamily of small leucine-rich proteoglycans (SLRPs) [10]. From a homology comparison of Nyctalopin with other SLRP proteins, it is evident that Nyctalopin is a unique member of this subfamily and the LRR superfamily in general. Nyctalopin has five putative consensus sequences (N-X-(S/T)) necessary for substitution by N-linked oligosaccharides or keratan sulfate [14], three of these sequences lie within the LRR region. The NH_2-terminal end of Nyctalopin is predicted [15] to contain a membrane signal peptide with a putative cleavage site between amino acid 23 and 24, AWA-VG (Figure 3). In addition, the carboxyl-terminal region of Nyctalopin contains a GPI-anchor signal sequence, including the requisite GPI N-terminal signal sequence (amino acids 339 to 379), the C-terminal hydrophobic region (last 22 amino acids) and a potential cleavage site at amino acids 445-447 [16] (Fig. 3b). The identification of these sites was accomplished at the website www.expasy.ch/tools, and is well known to those skilled in the art. Thus, *NALPN* appears to code for a GPI-anchored proteoglycan with a putative membrane signal peptide. Without being limited to a theory, these results suggest that the clinical features of complete X-linked CSNB can be explained by the presence of a mutant Nyctalopin causing the disruption of selected connections or interactions between retinal neurons, including those of the retinal ON-bipolar pathway.

It is understood that, because of genetic redundancy, the nucleotide sequence of *NALPN*

disclosed herein may be modified by making variations in sequence that do not alter the amino acid sequence of the resultant protein. The nucleotide sequence may also be modified to make conservative amino acid substitutions to the resultant protein, which do not alter, or do not significantly alter, the biological activity of the resulting molecule. The resulting modified nucleotide sequences are contemplated herein.

It is understood that the amino acid sequence of Nyctalopin disclosed herein may be modified by making minor variations in sequence, such as conservative amino acid substitutions or minor deletions or insertions that do not alter the activity of the protein, and the resulting modified proteins are contemplated herein. Suitable conservative substitutions of amino acids are known to those of skill in this art, and may be made generally without altering the biological activity of the resulting molecule. Such substitutions may also be made empirically.

The protein may be purified from a recombinant expression system. For protein expression, eukaryotic or prokaryotic expression systems may be generated in which the *NALPN* gene sequence, cDNA or genomic, is introduced into a plasmid or other expression vector which is then introduced into living cells. Constructs in which the *NALPN* cDNA sequence containing the entire open reading frame inserted in the correct orientation into an expression plasmid may be used for protein expression. Alternatively, portions of the normal or mutant *NALPN* sequences may be inserted.

Typical expression vectors contain promoters that direct the synthesis of large amounts of mRNA corresponding to the gene. They may also include sequences allowing for their

autonomous replication within the host organism, sequences that encode genetic traits that allow cells containing the vectors to be selected, and sequences that increase the efficiency with which the mRNA is translated. Stable long-term vectors may be maintained as freely replicating entities by using regulatory elements of viruses. Cell lines may also be produced which have integrated the vector into the genomic DNA and in this manner the gene product is produced on a continuous basis.

Expression of foreign sequences in bacteria such as *E. coli* require the insertion of the sequence into an expression vector, usually a plasmid which contains several elements such as sequences encoding a selectable marker that assures maintenance of the vector in the cell, a controllable transcriptional promoter which upon induction can produce large amounts of mRNA from the cloned gene, translational control sequences and a polylinker to simplify insertion of the gene in the correct orientation within the vector. A relatively simple *E. coli* expression system utilizes the lac promoter and a neighbouring lacZ gene which is cut out of the expression vector with restriction enzymes and replaced by the *NALPN* gene sequence.

In vitro expression of proteins encoded by cloned DNA is also possible using the T7 late-promoter expression system. Plasmid vectors containing late promoters and the corresponding RNA polymerases from related bacteriophages such as T3, T5 and SP6 may also be used for in vitro production of proteins from cloned DNA. E. coli can also be used for expression by infection with M13 Phage mGPI-2. *E. coli* vectors can also be used with phage lambda regulatory sequences, with fusion protein vectors, with maltose-binding protein fusions, and with glutathione-S-transferase fusion proteins.

Eukaryotic expression systems permit appropriate post-translational modifications of expressed proteins. This allows for studies of the *NALPN* gene and gene product including determination of proper expression and post-translational modifications for biological activity, identifying regulatory elements in the 5' region of the gene and their role in tissue regulation of protein expression. It also permits the production of large amounts of normal and mutant proteins for isolation and purification, the use of cells expressing *NALPN* as a functional assay system for antibodies generated against the protein, the testing of the effectiveness of pharmacological agents or to increase or decrease the activity of *NALPN*, and the study of the function of the normal complete protein, specific portions of the protein, or of naturally occurring polymorphisms and artificially produced mutated proteins.

In order to produce mutated or polymorphic proteins, the *NALPN* DNA sequence can be altered using procedures such as restriction enzyme digestion, DNA polymerase fill-in, exonuclease deletion, terminal deoxynucleotide transferase extension, ligation of synthetic or cloned DNA sequences and site-directed sequence alteration using specific oligonucleotides together with PCR. Alteration of the cDNA will allow for the production of specific mutations within cDNA sequence in order to express the created mutated proteins and study their biological effects.

Once an appropriate expression vector containing the *NALPN* gene is constructed, it is introduced into an appropriate *E. coli* strain by transformation techniques including calcium

phosphate transfection, DEAE-dextran transfection, electroporation, microinjection, protoplast fusion and liposome-mediated transfection.

The host cell to be transfected with the vector of this invention may be selected from the group consisting of *E.coli*, Pseudomonas, Bacillus-subtilis, or other bacilli or bacteria, yeast, fungi, insect (using baculoviral vectors for expression), mouse or other animal or human tissue cells. Mammalian cells can also be used to express the Nyctalopin using a vaccinia virus expression system.

Prokaryotic and eukaryotic expression systems allow various important functional domains of the protein to be recovered as fusion proteins and used for binding, structural and functional studies and also for the generation of appropriate antibodies.

In order to express and purify the protein as a fusion protein, the *NALPN* cDNA sequence is inserted into a vector which contains a nucleotide sequence encoding another peptide (eg. GST--glutathionine succinyl transferase). The fusion protein is expressed and recovered from prokaryotic (eg. bacterial or baculovirus) or eukaryotic cells. The fusion protein can then be purified by affinity chromatography based upon the fusion vector sequence and the Nyctalopin obtained by enzymatic cleavage of the fusion protein.

Fusion proteins are particularly advantageous because they provide a system for ensuring a good expression of the protein without making any alterations to the 5' end of the coding sequence or immediately preceding the start codon.

In the fusion approach, a cloned gene is introduced into an expression vector 3' to a carrier sequence coding for the amino terminus of a highly expressed protein. The carrier sequence provides the necessary signals for good expression and the expressed fusion protein contains a N terminal region encoded by the carrier.

Purified protein can also be used in further biochemical analyses to establish secondary and tertiary structure. The preparation of substantially purified Nyctalopin or fragments thereof allows for the determination of the protein tertiary structure by x-ray crystallography of crystal of Nyctalopin or by NMR. Determination of structure may aid in the design of pharmaceuticals to interact with the protein, alter protein charge configuration or charge interaction with other proteins, or to alter its function in the cell.

E. Mutation and Segregation Analysis of *NALPN*

To identify which mutations in *NALPN* cause complete X-linked CSNB, primers flanking and internal to each exon are used for direct DNA sequence analysis of the entire *NALPN* gene in affected individuals from families with complete X-linked CSNB. Primers are designed, genomic DNA is amplified and PCR products are purified and sequenced using methods known to those skilled in the art.

Once a DNA sequence change is identified by the mutation analysis, aside from DNA sequencing, segregation analysis may be accomplished by a number of techniques. Allele sizing, as described in [18] can be used to follow as little as a single base-pair insertion or deletion in a gene. Alternatively, the segregation of a larger insertion or deletion, such as the 24 bp deletion

mutation found in one patient with complete X-linked CSNB, can be followed by PCR analysis and gel electrophoresis. Primers are used to amplify by PCR, in affected, non-affected and carrier individuals, the region that encompasses the deletion (or insertion) of the nucleotide sequence. After electrophoresis of the amplified products on an agarose gel (or a polyacrylamide gel), the deletion (or insertion) is detected by the presence of a PCR fragment which is smaller (or larger) than the PCR fragment from a normal gene.

Alternatively, segregation analysis may be accomplished by following the loss or gain of restriction endonuclease recognition sites (restriction sites). Mutated and wild-type sequences are analyzed by a DNA analysis program, for example DNA Strider1.2™, looking for changes in DNA sequence that would result in a loss or gain of a restriction site. Once found, these changes can be used to track the mutation in families of affected individuals. Firstly, PCR is used to amplify the region of interest from the genomic DNA of affected, non-affected and carrier individuals. The PCR products are digested with the enzyme that will detect the mutation (in either a positive or negative sense). The digested products are electrophoresed through agarose and visualized to determine whether the restriction enzyme site is present or not, whichever the case may be, in the individual analysed.

Numerous additional methods for identifying mutations of *NALPN* in individuals, or tracing mutations of *NALPN* through families, including but not limited to Single Strand Conformational (SSC) and heteroduplex analysis, are obvious to one skilled in the art.

F. Possible consequences of *NALPN* mutations

Eleven different mutations have been identified in *NALPN*, none of which are observed in chromosomes from normal individuals. In Nyctalopin, there are 11 leucine-rich repeats, which are all highly conserved with respect to the consensus sequence in SLRPs, and these are flanked by cysteine clusters (see Figure 3(b)) [10]. The deletion of a portion of the cysteine cluster in the amino-terminal portion of Nyctalopin appears to be responsible for complete X-linked CSNB in six families, which highlights the importance of this conserved region. The mutation that causes a stop codon on the carboxyl-terminal side of the leucine-rich repeats and another cysteine cluster, likely affects the ability of the protein to anchor in the membrane, as the protein portion on the carboxyl-terminal side of this mutation is presumed to be important for GPI anchoring Nyctalopin in the cellular membrane. Mutations that replace a consensus amino acid with another amino acid are presumed to disrupt an essential amino acid function. Mutations that result in the insertion (deletion) of amino acids in the protein are presumed to alter the folding of the protein.

Missense mutations, and additions or deletions of amino acids are predicted to disrupt specific functions of intact Nyctalopin and therefore may be informative as to the structure-function relationship of the protein. Such information is presumed to be useful for targeting therapy for retinal disease, either by direct action on Nyctalopin or, indirectly, on proteins which interact with Nyctalopin.

G. Construction of Full-Length cDNA Clones

Full length cDNA clones are constructed by a variety of methods known to those skilled in the art. Such methods include screening a cDNA library with a labeled DNA probe of the gene of interest, identifying overlapping cDNA clones and ligating them together into one clone that contains the entire coding region. One can also obtain a full length cDNA clone in one step from a library, obviating the need to perform intermediary ligation steps. If the 5'or 3' end only of the clone is missing, methods such as RACE (rapid amplification of cDNA ends) are used to complete the sequence, or if the full length sequence is known, PCR amplification and ligation of the fragments onto the ends of the cDNA clone may be used. The cDNA product is cloned into the vector of choice, such as a pUC vector or pBR322.

Alternatively, PCR is used to amplify the gene from a cDNA library or a cDNA preparation using a Taq Polymerase that is designed for long range PCR, such as Pfu™ or Vent™ Polymerase, and the PCR fragment is ligated into a the vector of choice. A preferred vector is PCR2.1-TOPO™ (Invitrogen). The PCR primer at the 5' end (the forward primer) is designed to contain a small ribosomal binding site. The forward primer and the PCR primer at the 3' end (the reverse primer) also contain recognition sites for extremely rare cutting restriction endonucleases, such as *Not*I, for cloning into the vector. If the entire cDNA sequence cannot be obtained in one step, then overlapping PCR fragments can be amplified and ligated together into a vector using methods known to those skilled in the art. PCR products are verified by DNA sequencing and restriction digestion, to ensure that they are identical in sequence to the native cDNA. After ligation into the vector, the fragments are again checked for accuracy by restriction

analysis or DNA sequencing. A person skilled in the art may modify these methods as necessary, depending upon the exigencies presented in each particular step of the assembly.

H. Identification and Characterization of Murine *NALPN*

The knowledge of the sequence of the human *NALPN* gene can be used to identify and isolate the homologous gene in other mammalian species. Mouse retinal cDNA can be amplified by PCR with the primers used to amplify the human *NALPN* gene or with other primers, such as degenerate primers, that are designed by reference to the human *NALPN* sequence. Amplified PCR fragments, which are similar in size as the human PCR products, are sequenced and compared to the human *NALPN* sequence. Any fragments with substantial homology to the human sequence are presumed to be portions of the murine *NALPN* gene.

To obtain the remainder of the murine *NALPN* sequence, mouse specific primers sets can be designed from the mouse sequence known to that point. These primers can be used to amplify the additional regions of murine *NALPN*.

Finally, the 5' and 3' ends of the murine cDNA sequence for *NALPN* can be obtained by 5' and 3' RACE, using the Marathon™ cDNA Amplification Kit (Clontech). These methods are well known to those skilled in the art.

EXAMPLES

The following examples are intended to illustrate but not limit the invention. While they are typical of those that might be used, other procedures known to those skilled in the art may alternatively be utilized.

Example 1- Identification of the Genomic Region containing the CSNB1 locus

Twenty-four families with complete X-linked CSNB were included in this study. The diagnosis of complete X-linked CSNB in these families involved electrophysiological and psychophysical testing [18], which established the reduction or alteration of the rod pathway- and cone pathway-mediated function in the retina of these patients. The results of this type of testing on normal and affected subjects is shown in Figure 1. Panel "a" shows that the virtual absence of a rod response, the relative preservation of the scotopic white-flash a-wave with a severely subnormal b-wave, and the loss of the first two major photopic oscillatory potentials for the photopic single flash. Panel "b" shows the loss of rod sensitivity across the retina, and scattered loss of cone sensitivity. Panel "c" the spectral sensitivity measurements, dark adapted, in a normal subject and a patient.

The preliminary genotype analysis of our families was performed as described [6,7]. Three new markers were developed based on dinucleotide repeats that were identified in large-scale DNA sequence. Primers pairs for these markers are as follows:

TABLE 1

Polymorphism	Forward Primer	Reverse Primer
506B13CA1	atcacagtgccctgcctaaa (SEQ. ID.No.3)	tcccaaagtgctgggattac (SEQ. ID.No.4)
200L4CA1	gaacagcaaaccaaatccaaa (SEQ. ID No. 5)	ggcctatggtaatgcctcct (SEQ. ID No. 6)
169I5CA2	aaacttagctgggcatgctg (SEQ. ID No. 7)	gctgggactacatacagcaca (SEQ. ID No. 8)

Using these markers and other know markers, an analysis of selected recombinant X chromosomes in the set of families with complete X-linked CSNB enable us to moved the distal boundary of the CSNB1 minimal region from DXS556 to the interval between 200L4CA1 and DXS8012 (Fig. 2 and Fig. 3(a)). From further analysis, the proximal crossover previously observed in patient V:1 in family P23 [19] was limited to the interval between DXS1207 and DXS228 (Fig. 2 and Fig. 3(a)).

Example 2 – Sequence of the full length *NALPN* cDNA

To position genetic markers accurately across the CSNB1 minimal region and identify candidate genes for the CSNB1 locus, a robust physical map of the CSNB1 minimal region in Xp11.4 was developed. A subset of BAC and PAC clones from the minimal tiling path of the estimated 1.2 Mb CSNB1 minimal region is shown in Figure 3. BAC clones 378P5, 36P21, 160H17, and 317C4, which had not been sequenced at the Sanger Center, were sequenced to between 1-2.7-fold redundancy to identify additional candidate genes. A sub-library was constructed [20] for each of the BAC clones. DNA from each BAC was isolated, randomly

sheared by nebulization, and fractionated by agarose gel electrophoresis. Fragments (2-4 kb) were collected, blunt-ended, and cloned into M13mp19 using standard techniques. Random clones from each sub-library were sequenced with the aid of ABI 373 or 377 sequencing machines and fluorescently labeled primers (ABI, Amersham). DNAStar software was used for gel trace analysis and contig assembly as well as DNA and protein alignments. DNA and protein sequences were examined against available public databases using the various Blast programs available through the network server at the National Center for Biotechnology Information.

The extended *NALPN* cDNA sequence was established by sequencing of PCR and RACE products. First-strand cDNA from total human retinal RNA was used as the template for PCR. Touchdown PCR using the Failsafe™ PCR Premix Selection Kit was carried out according to the manufacturer's protocol (Epicentre Technologies). RACE was carried out using Human Retina Marathon-ready cDNA (Clontech). As the GC content of exon 3 of *NALPN* averages 72%, Advantage-GC 2 Polymerase (Clontech) was used, and touchdown PCR for 5' RACE (94°C for 30s, 72°C for 4 min. for 5 cycles; 94°C for 30s, 70°C for 4 min. for 5 cycles; 94°C for 30s, 68°C for 4 min. for a further 25 cycles). A secondary amplification using a nested *NALPN* specific primer was then performed using the same conditions as above. RACE and PCR products were gel purified using the Concert Gel Extraction kit (Life Technologies) and sequenced using the ThermoSequenase ^{33}P-radiolabelled terminator cycle sequencing kit (Amersham Life Science). To establish the genomic organization of the *NALPN* gene, the full-length cDNA sequence of *NALPN* was compared to the genomic sequence derived from our analysis of BAC clone 378P5 and that produced by the Sanger Centre for BAC clone 169I5.

BAC clone 378P5 (Fig. 3(a)) yielded a sequence that had partial complete homology with a 526-bp expressed sequence tag Q14392 (Accession No. AI861796). Sequence of the BAC clone 378P5 in the region of homology to ESTQ14392 overlaps with the partial DNA sequence from BAC clone 169I5(Fig. 3(a)). GenScan and GeneFinder analysis of a 20 kb portion of the genomic sequence from clone 169I5 (Sanger Centre, www.sanger.ac.uk) that encompasses EST Q14392 predicted a novel open reading frame which we have designated *NALPN* and is shown in Firgure 3(a).

Example 3- Expression of *NALPN* in Humans

Expression of *NALPN* was assessed by PCR amplification of a QUICK-Screen™ Human cDNA Library Panel, (Clontech) using primers spanning exon 1 to 3:

GXC1NF (Forward)	5'-AGGGAGTGGAGGGGACCTCAG-3'; (SEQ. ID. No. 9)
GXC1N3R (Reverse)	5'-ACGGCACGGACGCGGTTG-3' (SEQ. ID No. 10)

These primers generate a 755-bp product using buffer K from the Failsafe™ PCR Premix Selection Kit (Epicentre Technologies), and touchdown PCR (94°C for 1min., 65°C to 55°C over 10 cycles for 30 sec., 72°C for 1min. 30 sec.; then 94°C for 1min, 55°C for 30 sec., 72°C for 1min 30 sec. for a further 29 cycles). PCR products were electrophoresed through a 1% agarose gel and visualized by ethidium bromide staining. The ubiquitously expressed EST JRL4A1 was used as an amplification control. The 755-bp PCR product was detected in the retinal and kidney cDNA library (Figure 4).

The RNA *in situ* hybridization method was performed as described [21-23]. Sections were hybridized with 668-bp antisense probe (from nt1557-2224 of the *NALPN* cDNA), at a concentration of 400 ng/ml. The hybridization was done in 50% formamide, 5 x sodium chloride-sodium citrate-phosphate (SSCP) and 40 mg/ml salmon sperm DNA, for 18 hr at 65°C under parafilm (American National Can, Chicago, IL). Two post-hybridization washes were performed with 2 x SSCP for a total of 30 min at 68°C; followed by two washes in 0.1 x SSCP for a total of 1 hr at 68°C. Incubation with Fab fragments from an anti-digoxigenin antibody (1:5000) from sheep, conjugated with alkaline phosphatase (Boehringer Mannheim) for 2 hr at 22°C, followed by washes with SSCP, was used for the detection of the digoxigenin-labeled riboprobes. Precipitation of the reaction products of nitroblue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate (Life Technologies) continued for 3–5 days with a daily change of substrate solution. Slides were mounted with Glycergel (Dako, Mississauga, Ontario, Canada). Images were obtained digitally on a Carl Zeiss Axioskop 2 microscope with a cooled CCD camera (Diagnostic Instruments, Inc.). The images were converted to a grey scale, and the brightness and contrast were adjusted in Adobe Photoshop.

Example 4: Mutation and segregation Analysis of *NALPN*

Mutations were initially identified by direct DNA sequencing of PCR amplified exons from males affected with complete X-linked CSNB. PCR primers were designed to amplify from adjacent intron sequences 20-50 bp from the splice site. Additional sequencing primers internal to each exon were also used to establish total DNA sequence across an exon. A minimum of 90 control chromosomes from unrelated Caucasians were evaluated using the same

techniques to test for nucleotide changes in *NALPN*.

Genomic DNA (100-200 ng) was amplified with Platinum™ Taq polymerase (Life Technologies), in 1.5-2.5 mM $MgCl_2$, and 6% DMSO, in the supplied buffer. Touchdown PCR conditions were the same as described for establishing the nucleotide sequence of *NALPN* (see Example 2). Purified PCR products from affected and control individuals were sequenced using ThermoSequenase™ radiolabeled terminator cycle sequencing (Amersham LIFE SCIENCE), electrophoresed on polyacrylamide gels, and visualized by autoradiography.

This analysis revealed a total of 11 different sequence alterations in *NALPN* from 17 of the families that were studied. These mutations are summarized in Figure 6. The observed sequence alterations segregated within families to each of the affected members and through carrier females. None of the mutations identified in patients with complete X-linked CSNB was observed in 90 control chromosomes.

One mutation (Figure 5A), that changed a tryptophan codon to a stop codon, was detected in patients from two large pedigrees of Hispanic origin from Costa Rica. This mutation, (designated W350X in Figure 6) occurs on the carboxyl side of the leucine-rich repeats and cysteine clusters of Nyctalopin, and is shown in Figure 5(a).

Six other families were found to have an in-phase 24 nt deletion that results in the loss of eight amino acids - RACPAACA (see Figure 5(b)). Six of these amino acids form part of a conserved cysteine-cluster on the amino-terminal side of the leucine-rich repeats, as shown in

Figure 3(B). Haplotype analysis of X chromosomes with this deletion mutation from each of the six families revealed nearly identical haplotypes, suggesting that these families share a common founder mutation. In three families, insertion mutations representing duplications of adjacent protein sequence add either six or three amino acids (Figure 3(b)).

Single-nucleotide changes in six other families with complete X-linked CSNB are predicted to cause missense mutations. Four of these mutations change a conserved leucine to a glutamine or proline; one changes a conserved Asparagine to a Lysine and another changes a conserved Phenylalanine to a Serine (Figure 6). Figure 5(d) demonstrates one of these particular missense mutations, a T to an A which results in a Leucine to Glutamine change.

Once a nucleotide change was identified, the loss or gain of restriction sites of the PCR fragment was analyzed using DNA Strider 1.2™. For example, the G to A transition in the *NALPN* gene in the W350X mutation destroys a *Fok*I site. This change in restriction enzyme sites can be used for segregation analysis as shown in Figure 5(a). PCR primers that will amplify over the region where the mutation occurs are designed as described above. The primers used were:

Forward: GATTTTTCCTGGGGTGACCT (SEQ. ID. No. 11)

Reverse: GTCCAGGTCGATGGAGACC (SEQ. ID. No. 12)

Genomic DNA samples from affected, non-affected and carrier individuals were amplified by PCR, as described above and digested with *Fok*I. The products were separated on an agarose gel and visualized by ethidium bromide staining. In the normal *NALPN* gene, the 916 bp PCR fragment which amplifies the region where the W350X mutation would be located, is

digested by the restriction endonuclease *FokI* into 568- and 348-bp fragments. The PCR fragment from chromosomes of affected individuals is not digested by *FokI*. The PCR fragments from carrier individuals exhibit all three bands, the PCR product from the mutated gene , and the two digested bands from the normal gene.

Segregation analysis can also be performed, as shown in Figure 5(c). Primers were designed as described above to amplify over the region where the 8-amino acid deletion occurs. The primers used were:

Forward: GATTTTTCCTGGGGTGACCT (SEQ. ID. No. 13)

Reverse: GTCCAGGTCGATGGAGACC (SEQ. ID. No. 14)

PCR was used to amplify the region of interest from genomic DNA of affected, non-affected and carrier individuals. The PCR products were electrophoresed through agarose gels. In Figure 5(C) the gene carrying the deletion was identified in affected and carrier individuals by a 238 bp PCR fragment. The PCR fragment from normal, non-deleted chromosomes was 262 bp.

WHAT IS CLAIMED IS:

1. An isolated DNA molecule encoding a mammalian retinal and kidney GPI-anchored small leucine-rich proteoglycan.

2. The isolated DNA molecule of claim 1 wherein said DNA is cDNA.

3. The isolated DNA molecule of claim 1 wherein said DNA is human DNA.

4. The isolated DNA molecule of claim 1 wherein said DNA is murine DNA.

5. The isolated DNA molecule of claim 1 wherein said DNA encodes an amino acid sequence which is at least 50% identical to SEQ ID NO 2.

6. The isolated DNA molecule of claim 1 wherein said DNA encodes the amino acid sequence of SEQ ID NO 2 or SEQ ID NO 2 with conservative amino acid substitutions.

7. The isolated DNA molecule of claim 1 wherein said DNA has the nucleotide sequence corresponding to SEQ ID NO 1.

8. An isolated DNA molecule comprising a sequence that hybridizes under stringent conditions to a hybridization probe having a nucleotide sequence of SEQ ID NO 1 or the complement of SEQ ID NO 1.

9. An expression vector comprising a DNA sequence encoding a mammalian retinal and kidney GPI-anchored small leucine-rich proteoglycan.

10. A cultured cell comprising the expression vector of claim 9.

11. A cultured cell comprising the DNA sequence of claim 1, operably linked to an expression control sequence.

12. A cultured cell transfected with the vector of claim 9, or a progeny of said cell, wherein the cell expresses the proteoglycan.

13. A method of producing a proteoglycan, comprising culturing the cell of claim 10, 11 or 12 under conditions permitting the expression of the proteoglycan.

14. The method of claim 13 further comprising the step of purifying the proteoglycan from the cell or the medium of the cell.

15. A purified polypeptide having an amino acid sequence comprising one of:

(a) SEQ ID NO 2;

(b) SEQ ID NO 2 having at least one conservative amino acid substitition; or

(c) an amino acid sequence which is at least 50% identical to SEQ ID NO 2.

16. A purified mammalian retinal and kidney GPI-anchored small leucine-rich proteoglycan

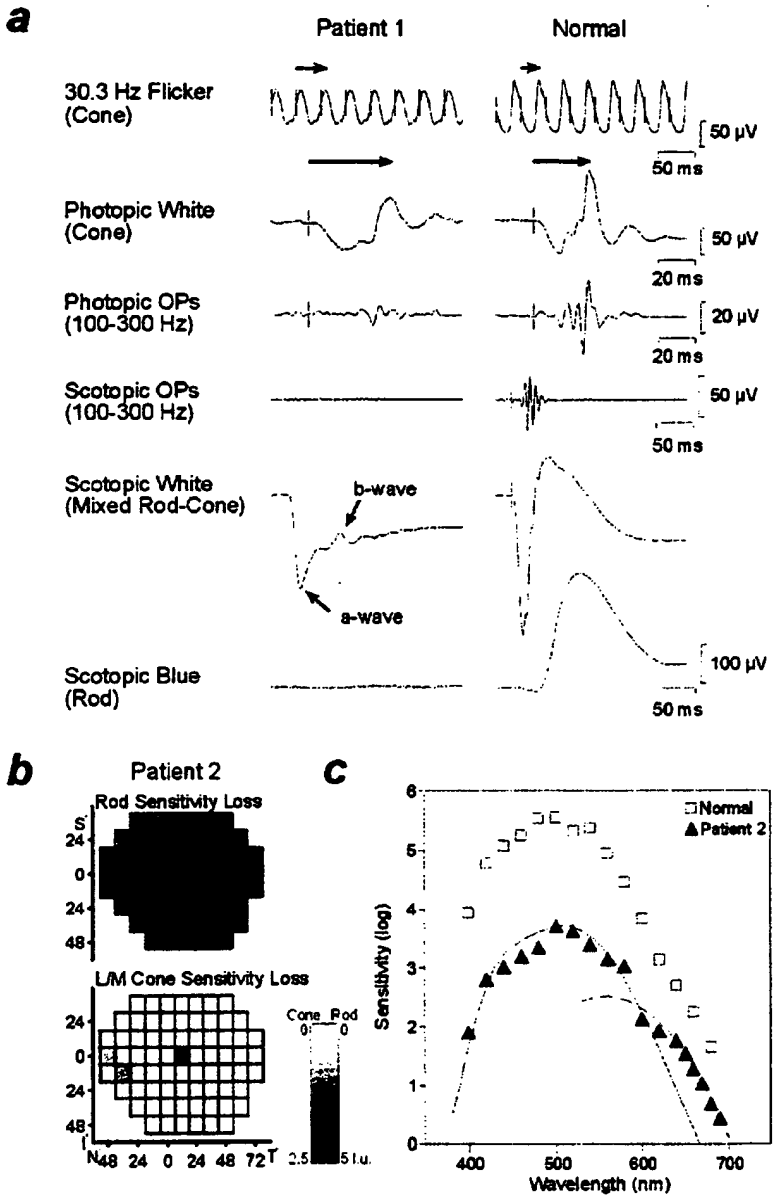

a
Patient 1
Normal
30.3 Hz Flicker (Cone)
50 µV
50 ms
Photopic White (Cone)
50 µV
20 ms
Photopic OPs (100-300 Hz)
20 µV
20 ms
Scotopic OPs (100-300 Hz)
50 µV
50 ms
Scotopic White (Mixed Rod-Cone)
b-wave
a-wave
Scotopic Blue (Rod)
100 µV
50 ms
b
Patient 2
Rod Sensitivity Loss
L/M Cone Sensitivity Loss
Cone Rod
0.5 5 l.u.
c
Normal
Patient 2
Sensitivity (log)
Wavelength (nm)

FIGURE 1

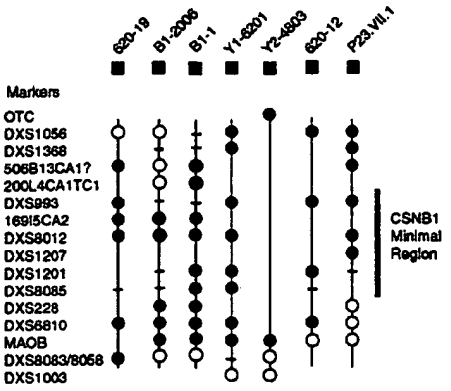
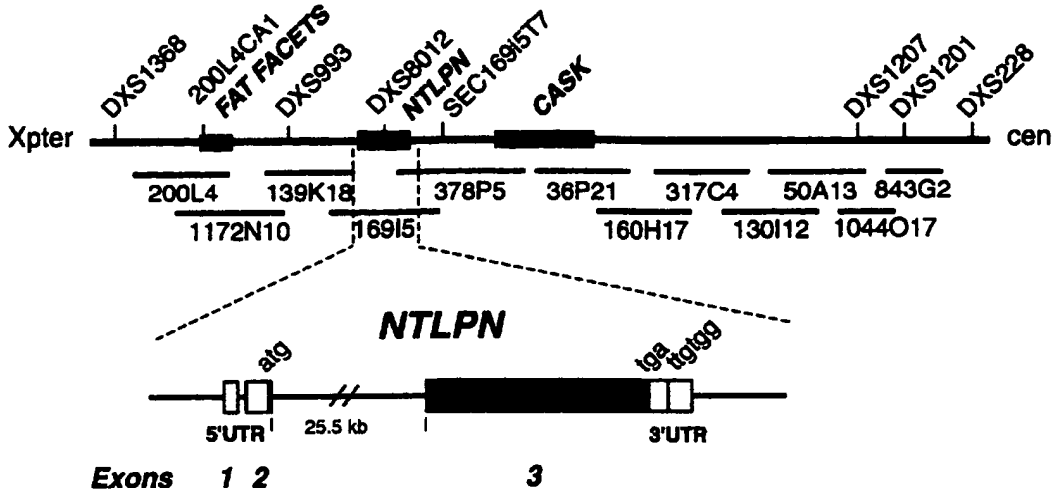

620-19
B1-2006
B1-1
Y1-6201
Y2-4803
620-12
P23.VII.1
Markers
OTC
DXS1056
DXS1368
506B13CA17
200L4CA1TC1
DXS993
16915CA2
DXS8012
DXS1207
DXS1201
DXS8085
DXS228
DXS6810
MAOB
DXS8083/8058
DXS1003
CSNB1
Minimal
Region

FIGURE 2

a Physical map of the CSNB1 minimal region and genomic rganization of *NTLPN*



b Leucine-rich repeats and the distribution of mutations in the Nyctalopin protein

```
MKGRGMLVLLLHAVVLGLPSAWAVGA ARA PAA A STVERG SVR DR  - 50
                           del
1.    AGLLRVPAELPCEAVSIDLDRNGL                      - 74
2.    RFLGERAFGTLPSLRRLSLRHNNL                      - 98
3.    SFITPGAFKGLPRLAELRLAHNGDL                     - 123
4.    RYLHARTFAALSRLRRLDLAACRL                      - 147
5.    FSVPERLLAELPALRELAAFDNLFRRVPGALRGL            - 181
6.    ANLTHAHL15LRRLRSLSLQANRV                      - 218
       ERGRIEAVASSSLQG
7.    RAVHAGAFGDCGVLEHLLLNDNLL                      - 242
8.    AELPADAFRGLRRLRTLNLGGNAL                      - 266
9.    DRVARAWFADLAELELLYLDRNSI                      - 290
10.   AFVEEGAFQNLSGLLALHLNGNRL                      - 314
11.   TVLAWVAFQPGFFLGRLFLFRNPW                      - 338
        β-sheet     α-helix
                |
 D RLEWLRDWMEGSGRVTDVP ASPGSVAGLDLSQVTFGRSSDGL VD  - 388
PEELNLTTSSPGPSPEPAATTVSRFSSLLSKLLAPRVPVEEAANTTGGLA - 438
NASLSDSLSSRGVGGAGRQPWFLLASCLLPSVAQHVVFGLQMD       - 481
```

^ *insertions of SVPERLL, GLR and RLR, respectively*
v *most likely signal peptide cleavage site*

FIGURE 3

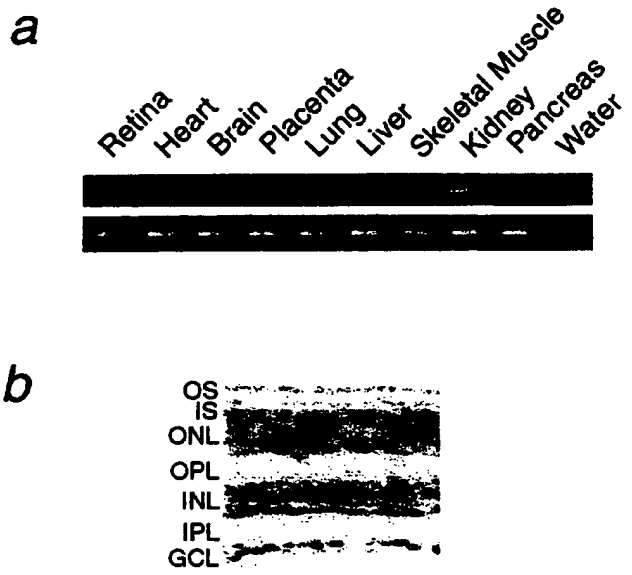

a
Retina
Heart
Brain
Placenta
Lung
Liver
Skeletal Muscle
Kidney
Pancreas
Water
b
OS
IS
ONL
OPL
INL
IPL
GCL

FIGURE 4

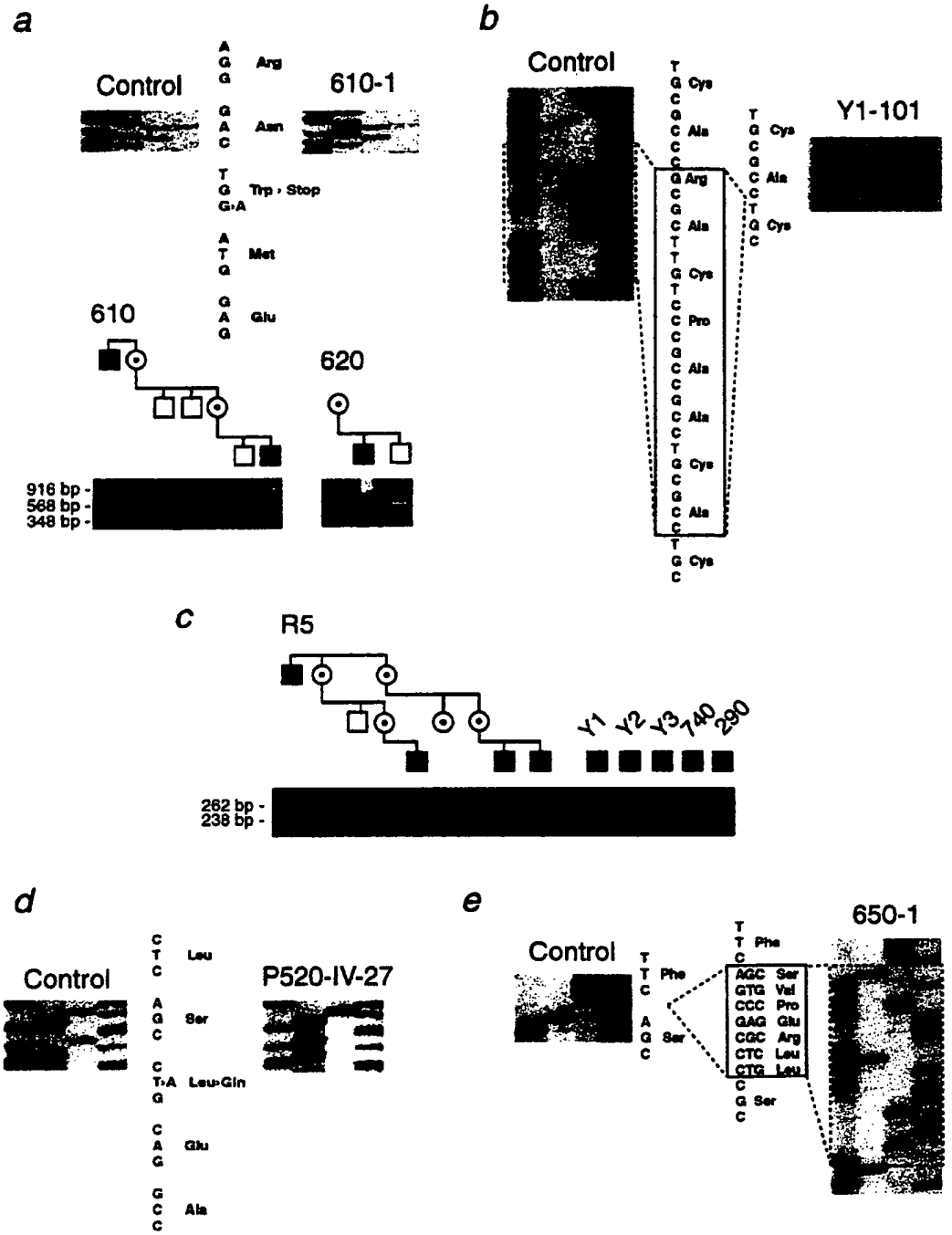

a
Control
610-1
Arg
Asn
Trp → Stop
Met
Glu
610
620
916 bp -
568 bp -
348 bp -
b
Control
Y1-101
Cys
Ala
Arg
Ala
Cys
Pro
Ala
Ala
Cys
Ala
Cys
c
R5
Y1
Y2
Y3
740
290
262 bp -
238 bp -
d
Control
P520-IV-27
Leu
Ser
Leu→Gln
Glu
Ala
e
Control
650-1
Phe
Ser
Ser
Val
Pro
Glu
Arg
Leu
Leu
Ser

FIGURE 5

Table 1 • Nyctalopin mutations in families with complete CSNB

Family	Origin	Mutation	Codon change	Predicted effect on Nyctalopin protein
Y1, Y2, Y3, R5, 290, and 740	USA	85-108del24nt	RACPAACA29-36del	Loss of part of NH_2 Cys cluster
650	Canada	444-445ins21nt	SVPERLL148-149ins	Expansion between LRR4 and LRR5
750	Canada	551T→C	L184P	Missense, Leucine to Proline
640	USA	610-611ins9nt	GLR203-204ins	Expansion of loop in LRR6
R7	USA	615-616ins9nt	RLR205-206ins	Distortion of LRR6 β-sheet to α-helix transition
P520 (890)	NL	638T→A	L213Q	Missense, Leucine to Glutamine
550	Germany	695T→C	L232P	Missense, Leucine to Proline
B1	USA	792C→G	N264K	Missense, Asparagine to Lysine
B660	USA	854T→C	L285P	Missense, Leucine to Proline
B2	USA	893T→C	F298S	Missense, Phenylalanine to Serine
610, 620	Costa Rica	1049G→A	W350X	Protein truncation, loss of GPI-anchoring signals

FIGURE 6

```
          -98                                                                                                  -1
          Ggctgagggagtggaggggacctcagaggagcaggaccagggagactcccaggacggtaggggtcccacggctgggtggtcctaagccactgggtgg

   1/1    ATGaaaggccgagggatgttggtcctgcttctgcatgcggtggtcctcggcctgcccagcgcctgggccgtgggggcctgcgcccgcgcttgtcccgccgcctgcgcctgcagcaccgtg
          M  K  G  R  G  M  L  V  L  L  L  H  A  V  V  L  G  L  P  S  A  W  A  V  G  A  C  A  R  A  C  P  A  A  C  A  C  S  T  V

 121/41   gagcgcggctgctcggtgcgctgcgaccgcgcgggcctcctgcgggtgccggccgagctcccgtgcgaggcggtctccatcgacctggaccggaacggcctgcgcttcctgggcgagcga
          E  R  G  C  S  V  R  C  D  R  A  G  L  L  R  V  P  A  E  L  P  C  E  A  V  S  I  D  L  D  R  N  G  L  R  F  L  G  E  R

 241/81   gccttcggcacgctgccgtccttgcgccgcctgtcgctgcgccacaacaacctgtccttcatcacgcccggcgccttcaagggcctgccgcgcctggctgagctgcgcctggcgcacaac
          A  F  G  T  L  P  S  L  R  R  L  S  L  R  H  N  N  L  S  F  I  T  P  G  A  F  K  G  L  P  R  L  A  E  L  R  L  A  H  N

 361/121  ggcgacctgcgctacctgcacgcgcgcaccttcgcggcgctcagccgcctgcgccgcctagacctagcagcctgccgcctcttcagcgtgcccgagcgcctcctggccgaactgccggcc
          G  D  L  R  Y  L  H  A  R  T  F  A  A  L  S  R  L  R  R  L  D  L  A  A  C  R  L  F  S  V  P  E  R  L  L  A  E  L  P  A

 481/161  ctgcgcgaactcgccgccttcgacaacctgttccgccgcgtgccgggcgcgctgcgcggcctggccaacctgacgcacgcgcacctggagcgcggccgcatcgaggcggtggcctccagc
          L  R  E  L  A  A  F  D  N  L  F  R  R  V  P  G  A  L  R  G  L  A  N  L  T  H  A  H  L  E  R  G  R  I  E  A  V  A  S  S

 601/201  tcgctgcagggcctgcgccgcctgcgctcgctcagcctgcaggccaaccgcgtccgtgccgtgcacgctggcgccttcggggactgtggcgtcctggagcatctgctgctcaacgacaac
          S  L  Q  G  L  R  R  L  R  S  L  S  L  Q  A  N  R  V  R  A  V  H  A  G  A  F  G  D  C  G  V  L  E  H  L  L  L  N  D  N

 721/241  ctgctggccgagctcccggccgacgccttccgcggcctgcggcgcctgcgcacgctcaacctgggtggcaacgcgctggaccgcgtggcgcgcgcctggttcgctgacctggccgagctc
          L  L  A  E  L  P  A  D  A  F  R  G  L  R  R  L  R  T  L  N  L  G  G  N  A  L  D  R  V  A  R  A  W  F  A  D  L  A  E  L

 841/281  gagctgctctacctggaccgcaacagcatcgccttcgtggaggagggcgccttccagaacctctcgggtctcctcgcgctgcacctcaacggcaaccgcctcaccgtgctcgcctgggtc
          E  L  L  Y  L  D  R  N  S  I  A  F  V  E  E  G  A  F  Q  N  L  S  G  L  L  A  L  H  L  N  G  N  R  L  T  V  L  A  W  V

 961/321  gccttccagcccggcttcttcctgggccgcctcttcctcttccgcaacccgtggtgctgcgactgccgtctggagtggctgagggactggatggagggctccggacgtgtcaccgacgtg
          A  F  Q  P  G  F  F  L  G  R  L  F  L  F  R  N  P  W  C  C  D  C  R  L  E  W  L  R  D  W  M  E  G  S  G  R  V  T  D  V

1081/361  ccgtgcgcctccccgggctccgtggccggcctggacctcagccaggtgaccttcgggcgctcctccgatggcctctgtgtggaccccgaggagctgaacctcaccacgtccagtccaggc
          P  C  A  S  P  G  S  V  A  G  L  D  L  S  Q  V  T  F  G  R  S  S  D  G  L  C  V  D  P  E  E  L  N  L  T  T  S  S  P  G

1201/401  ccgtccccagaaccagcggccaccaccgtgagcaggttcagcagcctcctctccaagctgctggccccgagggtcccggtggaggaggcggccaacaccactgggggctggccaacgcc
          P  S  P  E  P  A  A  T  T  V  S  R  F  S  S  L  L  S  K  L  L  A  P  R  V  P  V  E  E  A  A  N  T  T  G  G  L  A  N  A

1321/441  tccctgtccgacagcctctcctcccgtggggtgggaggcgcgggccggcagccctggtttctcctcgcctcttgtctcctgcccagcgtggcccagcacgtggtgtttggcctgcagatg
          S  L  S  D  S  L  S  S  R  G  V  G  G  A  G  R  Q  P  W  F  L  L  A  S  C  L  L  P  S  V  A  Q  H  V  V  F  G  L  Q  M

1441/481  gactgacctggccagaggggggaaagtttgcttaactgggcttgagtgtgtttgtggtaaggggagaggagccggaatggagggcagaggtgaaaatcccagtggagggtggaaggaaccc
          D  *

   1561   Gtttgcctccagagatggccccagggagaacacagggacgtgccactcgaggggggaggatggtatggatttctgcttttgtcacacgggcatccattggaaaagagaagcaagaatgaac
   1681   gtgggccctcgggtgggaagactaggaatcggaagcttctagggcttcacatcccttcccctccctcccctcccttccctcatcttccaggcaacagtgcctgcaaggcctgaattagagag
   1801   acttccattggctaagtagttaagagccgtcccatttctcctggcggggtaacccattacaccgaagtcctttgttttctaccacaatcctcctcctcctctccaggggcctggaaacac
   1921   taggattcaggaaggtaggcaggacgtgagagaagggagatgggagagagatttaagacaaagggtggcggtggttcctggggtctgagatgtgttaggaggcgtttaaaacaaagatcc
   2041   agttcatttactccacagttattcccagggctggccctagccacaaaggaacttttagggcagggtagggaaaaaaggggcagcaggggggtgtgtttgtggacaaataaatttgtaaagtc
   2161   cgaggattaaaaaaaaaaaaaggttaaaccggtttctctactgctggacttttcagagtctttc
```

FIGURE 7